Abstract

Statistical literacy and statistical reasoning have been considered by the statistics education community as important learning goals to be developed in introductory statistics courses (Garfield & Ben-Zvi, 2008). Many statistics educators and scholars have tried to define these learning goals (e.g., Gal, 2002; Watson & Callingham, 2003; Garfield, 2002; Garfield & Chance, 2000; Wild & Pfannkuch, 1999). However, there is a lack of agreement regarding these definitions and the relationship between statistical literacy and statistical reasoning. In addition, there are assumptions in the statistics education literature of an overlap between statistical literacy and statistical reasoning (e.g., Rumsey, 2002; Budgett & Pfannkuch, 2007, and delMas, 2002) and of a hierarchy between these learning goals (Garfield & Ben-Zvi, 2008 and delMas, 2002). Empirical evidence is needed to support these assumptions.

The purpose of this study was to investigate how statistical literacy and statistical reasoning are related. Specifically, this research aimed to verify if these two learning goals are distinct or if they overlap. The three research questions addressed by this study were (1) what measurement model best represents the construct of statistical literacy and the construct of statistical reasoning given the criteria of fit and parsimony? (2) what measurement model best represents the construct of statistical literacy and the construct of statistical reasoning given the criteria of reliability and distinction? (3) what measurement model is most useful for understanding the constructs of statistical literacy and statistical reasoning?

To answer the three research questions, the REALI instrument was developed to concurrently measure statistical literacy and statistical reasoning. This instrument is composed of 40 items with 20 items measuring statistical literacy and 20 item measuring statistical reasoning. The items in this instrument assess eight areas of learning: (1) representations of data, (2) measures of center, (3) measures of variability, (4) study design, (5) hypothesis testing and p-values, (6)

confidence intervals, (7) bivariate data, and (8) probability. During the development process, several types of validity evidence were gathered to support the intended inferences and uses of the REALI's scores (and subscores): expert reviews, response process interviews with students, a pilot test, a field test, reliability, and a psychometric analysis.

Data from the field test were analyzed under the classical test theory (CTT) and item response theory frameworks (IRT). Five IRT models were fitted to the data: A Unidimensional Model, three bi-dimensional models (Uncorrelated Model, Correlated Model, and Cross-loading Model), and a bi-factor model (Bi-factor Model A). The main difference between these models was whether or not the model allowed statistical literacy and statistical reasoning dimensions to correlate and if a third dimension (statistical knowledge) was included in the model. These models were compared at the item- and model-level and the best fitting models were used to evaluate the relationships between the statistical literacy subscore and the statistical reasoning subscore. Evidence was found that the statistical literacy and reasoning subscores from the Cross-loading Model could be measured reliably and distinctly. In addition, statistical evidence also supported that reporting both statistical literacy and statistical reasoning subscores provided more distinct information than only reporting a unique score for each student. Such findings bring valuable information to the field of statistics education and can be used to guide instruction in introductory statistics courses. In addition, the REALI instrument is a tool that can be used by researchers and instructors to investigate students' understanding of statistical concepts or to evaluate new curricula.